THOUGHTS ON RENORMALIZATION

By Bill Daniel

INTRODUCTION

On page 80 of *Deep Down Things*, Bruce Schumm describes renormalization as a way to save quantum field theories from being "pecked to death by these ephemeral matter-antimatter fluctuations" (referring to Fig. 4.8). His description of renormalization over the next few pages is well done, but, by necessity for a book like this, incomplete. Many of the members of our group have an understanding of physics that is at a deeper level than that of Schumm's audience, and I thought there might be some interest in a somewhat (but not very much) deeper discussion of this important topic. Hence, these notes. If you're not interested, feel free to ignore this.

As always, please be aware that I am far from an expert in these matters, and this paper represents just my current (and probably flawed) understanding. Any corrections or additions are most welcome.

CUTOFF REGULARIZATION

A few problems in quantum mechanics are solvable in closed form, but these are the exceptions. Models of actual experimental configurations are not typically solvable, so they must be approached "perturbatively." This means that the physical situation under study is assumed to differ only slightly from one that *is* solvable. The Hamiltonian, the quantum mechanical operator that expresses the energy of the system, is typically the operator whose eigenvalues we wish to find. In a perturbative solution, the eigenstates of the actual Hamiltonian are assumed to be superpositions of those of the solvable Hamiltonian. This perturbative approach works, often very well, if the model of physical system is close to the chosen solvable one, but it will never be exact.

If you think of the case of an electron in an electromagnetic field, for example, we can diagonalize (solve) the electron's Hamiltonian as well as the Hamiltonian for the *free* field, but the total Hamiltonian for the interaction is not just their sum – it also includes an interaction term. This term makes it impossible to solve the total Hamiltonian in closed form. The problem can still be treated perturbatively, though, if the interaction is weak.

In Quantum Field Theory (QFT), where we work with cases in which special relativity is important, we use the alternative Lagrangian formulation, because Lagrangians are Lorentz invariant, while Hamiltonians are not. The Lagrangian for an interaction is composed of a sum of terms that are themselves the product of a "coupling constant" and a factor that accounts for the many ways momentum can potentially flow in the interaction. Coupling constants are not really constant; they vary with energy scale. This behavior is called "running" and we will see its importance later.

Each of these 2-factor terms in the Lagrangian is expressed as an infinite power series in the coupling constant with each term in the series corresponding to one or more Feynman diagrams. This series, called the “Dyson series,” may absolutely diverge, “superficially” diverge, or, in rare cases, it may converge to a finite limit.

The factor in each term that accounts for momentum is an integral, typically one with an infinite upper integration limit, of a function of a *positive power* of momentum, and so, in itself, is infinite. To evaluate series convergence, we must first make this factor finite. We do so by imposing an arbitrary cutoff momentum (or equivalently a cutoff length) and integrate up to that finite limit. (There are other, less brute force ways to do this called “regularization schemes,” but the cutoff approach is the most intuitive and illustrates the difficulties inherent in all of them. It’s the only one I will describe here.)

The idea is to experimentally measure the coupling constant at various momenta and work backwards to figure out what the, so-called, “bare coupling constant” at an infinite momentum would need to be to give the observed values. This approach introduces another characteristic length, that associated with the energy scale of the measurement. It’s important to realize that this is a *different scale* – not that of the cutoff length. In fact, the experimental scale is usually orders of magnitude larger than the arbitrary cutoff length that we use to regularize the momentum integrals.

At this point we can take the limit as the cutoff length goes to zero. Doing that implies, though, that spacetime is continuous and that the theory holds at *any* length scale – both fairly dubious assumptions. If that limit exists, that’s encouraging, but remember this is just one factor in one term of the Lagrangian. Also, remember that this calculation is perturbative, not exact – and there’s no guarantee that a perturbative approach will match reality. We will look at a better way to handle cutoff in a minute.

In quantum electrodynamics (QED), the theory of the interaction of light with matter, the bare charge of the electron becomes infinite as the cutoff approaches zero. This made lots of people, including Dirac and Feynman, uncomfortable, but the prevailing approach today has more of an Alfred E. Newman, “What, me worry?” flavor.

We don’t measure the bare charge, the current reasoning goes, and we never will, since the electron is always interacting with many of the fields that permeate the vacuum, so why worry about it? If the limit of a theory as distance goes to zero exists, even if it’s infinite, we say the theory is “renormalizable.” So, QED is a renormalizable, but in this sense, not a finite theory. The theory is quite serviceable, but its infinitude almost certainly means that one or both of the assumptions about continuous spacetime or QED being correct at any length scale must be wrong.

Vanishing Bare Coupling Constants

Recall that the integral accounting for momentum and the coupling constant are both functions of the cutoff length. This is bad. We don’t want our theory to depend on a cutoff – especially an arbitrary one! We saw that taking the limit as the cutoff length goes to zero works, but that it relies on some questionable assumptions. A better way to solve the problem is by adding some extra terms, called “counterterms,” to the Lagrangian. (For the moment let’s not worry about the physical consequences of doing this. I promise we will look at that question later.) Making an inspired choice of these counterterms allows them to cancel the effect of the cutoff up to some order in the Dyson series.

At first sight, this might seem like a problematic approach. “What if,” you ask, “including those counterterms just creates *new* infinites that need to be canceled by other counterterms, and so on?” Very insightful of you, I must say! This could be a real problem (and it is for some theories), but, in a surprising number of cases it turns out that only a few counterterms (3 in QED!) are necessary to cancel *all* divergences and make a theory independent of the cutoff. Theories that require only a finite number of such additional terms are renormalizable.

Freeman Dyson realized that all of this could be captured in one simple criterion for renormalizability. If we work in natural units (with $\hbar$ and c equal to 1 and unitless), coupling constants can be expressed in units of length to some power. Dyson’s criterion is that if all the coupling constants in a theory have powers of length less than or equal to zero, then that theory is renormalizable.

Our most important field theories today, QED, the electro-weak theory, and QCD are all renormalizable by the Dyson criterion. But again, remember that these are also all *perturbative* theories with the caveats that come with that.

But we can’t just go willy-nilly adding counterterms to a Lagrangian. They will generate their own Feynman diagrams in the Dyson series, and will have corresponding physical effects.

It turns out (you’re just going to have to trust me on this one) that the counterterms we add to the QED Lagrangian (and those for many other theories) have, not an *infinite* bare coupling constant like the bare charge on the electron, but a *zero* one. That doesn’t mean that they have no effect, though. At higher energy scales, those terms with a zero *bare* coupling constant can still have a non-zero *physical* coupling constant and therefore have an impact on interactions. In these cases, we say that the particle (or the field excitation, to be more accurate) acquires a mass or charge from its field interactions.

The best analogy I’ve seen for this is that of a ping pong ball. If you drop a ping pong ball, it will fall downward under the effect of the earth’s gravitational field – it has a positive mass. If you release it under water, though, it will rise. Its interaction with the water has, in a sense, given it a *negative* gravitational mass. If you imagine the ping pong ball to have *zero* mass in air, this is analogous to a particle having a zero bare mass, but, when the particle is immersed in a field at a particular energy scale (like the ball in water), it acquires mass. Don’t get too wrapped up in this analogy. Like all analogies, it’s not quite right, but maybe it’s helpful to get a visual picture of what’s going on.

There is a subtilty here that’s often overlooked. Renormalization is sometimes described in a way that leaves people thinking of it as a perhaps mathematically suspicious way to get rid of infinities. It does do that, but the real function of renormalization is to take an abstract field theory and make it apply to the real world!

To see how this works, consider the force on an electron brought near another electron. Coulomb’s Law says it will experience a repulsive force, $F \propto \frac{q^2}{r^2}$, where q is the charge on one electron and r is their separation. Electrons are point particles, so r can be as close to zero as we want. At $r = 0$, though, the force would be infinite. In the real world, this can never happen, of course. Even if we could overcome an infinite force, the uncertainty principle will not allow the electrons' positions to be pinned down accurately enough to establish $r = 0$. Coulomb's Law is an abstraction. It can't describe reality exactly, and the infinity that comes out of it is telling us this.

Lagrangians that lead to infinities are fundamentally no different. Bare electrons are a fiction. They are no more real than a point electric charge, and the quantum field theories that follow from them are just as much abstractions as is Coulomb's Law. Also like Coulomb's Law, such quantum field theories are still useful, just don't push them too hard.

So that's where things stood in about 1970. Physicists could hardly believe their luck. Though there were still infinities lurking around, they were manageable, and renormalization had saved their best theories from being "pecked to death" intractable infinities. Bring out the party hats! Well, maybe not so fast…

Renormalization Group Flow

In the early 1970's, Ken Wilson realized that cutoff regularization was masking what was really going on. Remember that I said that the cutoff length is usually set much smaller than the experimental length scale. Wilson focused on how system parameters depend on the scale at which we actually *measure* them, as opposed to some arbitrary cutoff scale.

To better understand what he did, think of a vector whose components are the coupling constants of a theory. The abstract space in which this vector lives is the space of quantum field theories, each characterized by a collection of coupling constants – the coordinates of the vector that corresponds to the theory.

Because those coupling constants "run," the vector associated with a theory traces a trajectory through this space as the energy scale changes. That trajectory is known as the "renormalization group flow." Wilson and others showed that, for low energy theories, the renormalization group flow is limited to a subspace in which the non-renormalizable terms are insignificant. (Keep in mind that "low energy" includes all energies achievable at the LHC, *i.e*. $\lesssim 10^4$ GeV!)

This decoupling of low energy theories is what physicists have in mind when they speak of "effective" field theories. This is where physics was in 1970; all the well-known quantum field theories of the time were effective theories by default. Effective theories work, sometimes spectacularly well, like calculating the gyromagnetic ratio of the electron using QED, but they break down in extreme cases like those in which quantum gravity becomes important.

This is both a good and a bad thing. It's good because it means that we don't have to have a complete theory to do useful work in a low energy regime because the theory is an island unto itself – completely detached from theories that describe higher energy environments. But it also means that our low energy experiments can give us *no information* about the non-renormalizable elements that become important at higher energies.

As the energy scale *decreases*, the renormalization group vector falls into what chaos theorists call a "basin of attraction." Multiple trajectories (field theories) can converge on the same basin of attraction or even land on the same unchanging "infrared fixed point." We get no help from low energy experiments in deciding which high energy theories actually describe the world. This is why none of our current theories of quantum gravity can claim evidentiary preeminence over any other. String theory, for example, doesn't provide corrections our current, low energy, field theories in the way special relativity was necessary to account for discrepancies in Newtonian mechanics at velocities near light speed. At high energies, we are truly in *terra incognita*.

As the energy scale *increases*, terms whose Dyson exponent is negative have coupling constants that go to zero. These theories become "free" theories at high energies, or short distances; *i.e.*, they exhibit "asymptotic freedom." There are good reasons to think that QCD, the theory of the strong force, is asymptoticly free. Because of the enormous energies locked in nucleons, quarks feel no strong force at distances less than about 10^{-15} m, but quickly become strongly coupled at longer distances.

String theory, and all of the other quantum field theories that come from quantizing gravity, are non-renormalizable. This is due to the fact that dimensions of the gravitational coupling constant (Newton's constant, G) in natural units includes mass, not just length.

Recently, Steven Weinberg and others have proposed that, though such theories are not asymptotically free, they *may* be what's come to be called "asymptotically safe" (or, with the even more imposing label, "non-perturbatively renormalizable"). This mouthful of terminology just means that, while their coupling constants don't go to zero as they would if they were asymptotically free, they do have a finite limit, known as an "ultraviolet fixed point" and similar to the "infrared fixed point" I mentioned above for renormalizable theories. That limit is unknown and may be very large, but its finitude may make non-renormalizable theories tractable. This remains an open question.

I hope all this has been reasonably intelligible and of some interest. Much of it is just barely within (or even outside) my frontier of understanding, so please forgive anything I have made a muddle of or even gotten completely wrong. Writing up these sorts of things is how I organize my thoughts and identify areas that need more study. I rarely share these writings because they are raw and incomplete, but I made an exception this time because I thought there might be interest in taking the ideas about renormalization in *Deep Down Things* a bit further.

Bill